Empirical Bayes

An example from practice

Disclaimers

1. I don’t know much theory.

2. I will oversimplify / obfuscate many Google-specific details.

3. All data is (obviously) simulated.

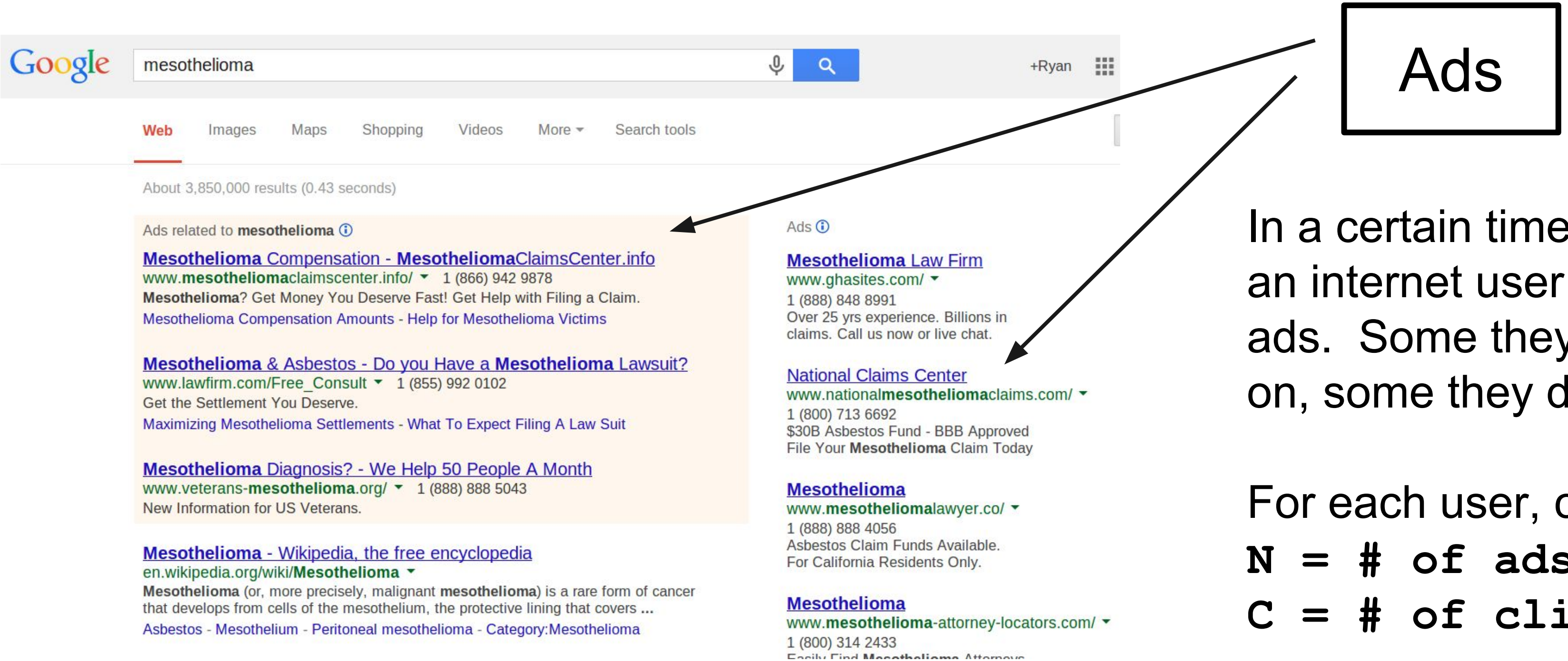
Questions to Keep in Mind

What is Bayesian about Empirical Bayes?

When can you use it?

How can you test its assumptions?

Framework



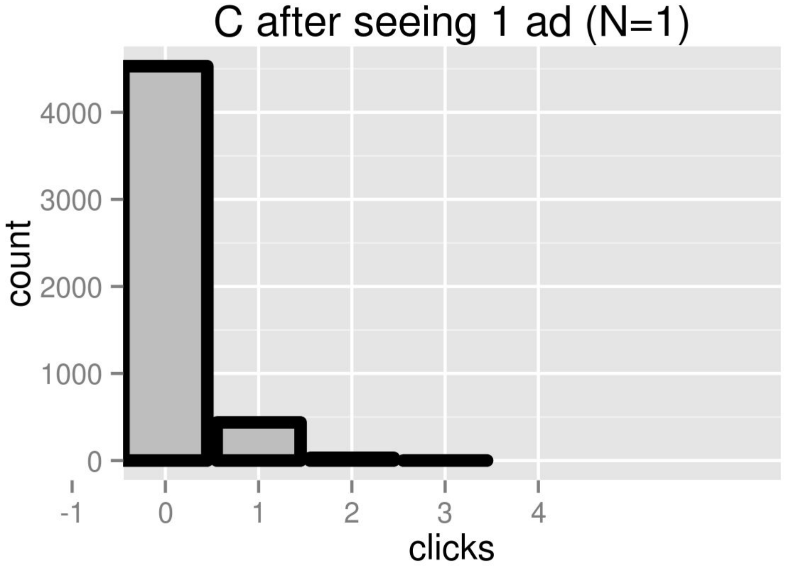
In a certain time period, an internet user sees N ads. Some they click on, some they don't.

For each user, count:

`N = # of ads`

`C = # of clicks`

A Dumb Question

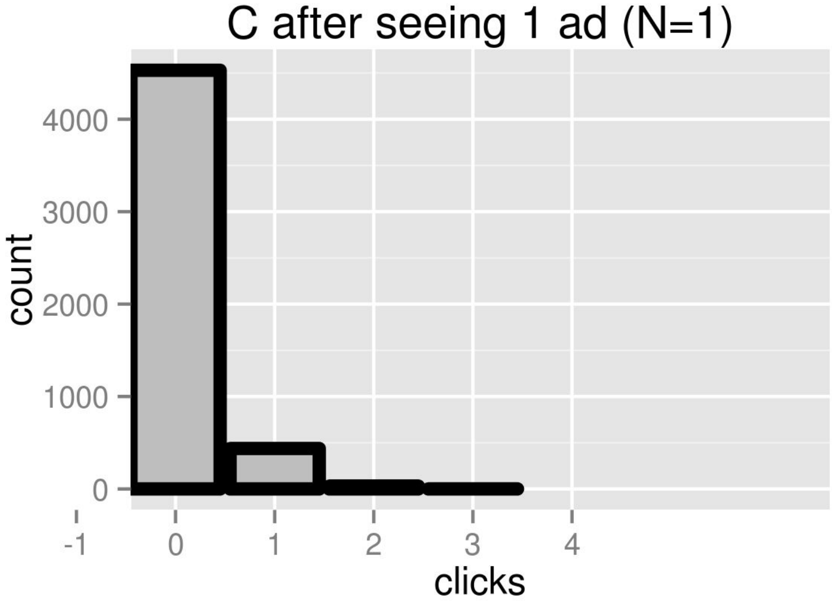


Look at 5000 people, each of whom saw one ad. The histogram of their clicks is shown.

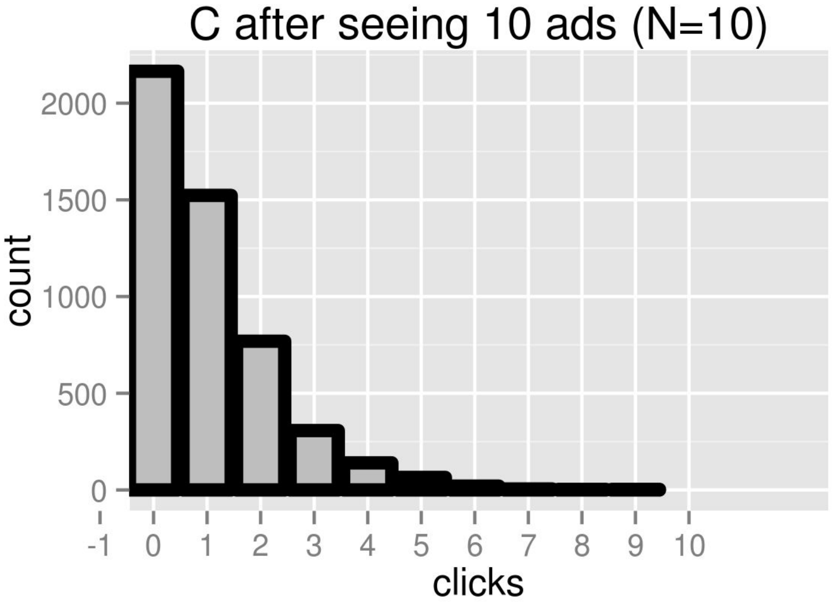
90.6% had no clicks.

Does this mean 90.6% of people never click on ads?

Same People, More Ads



90.6% no click



43.3% no click

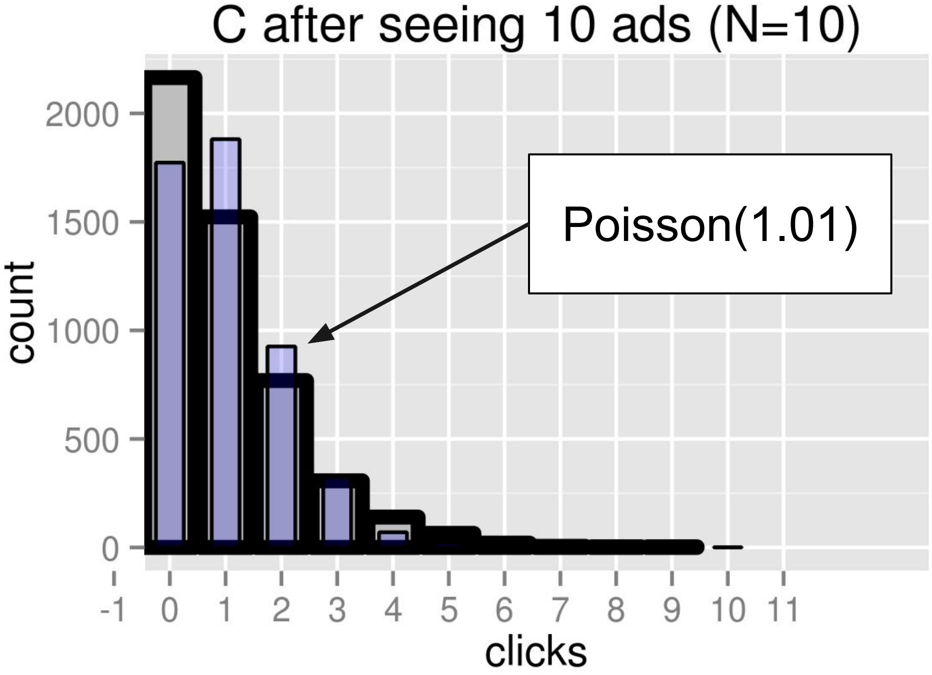
Clicks = Person + Noise

Clicks are a *noisy* observation of an attribute of a person at a particular time. Define:

p_i = the probability that person *i* clicks on an ad.

$(C|p_i) \sim Binom(N, p_i) \approx Poisson(Np_i)$

Overdispersion



Are all the p_i the same?

If so, because C is Poisson,

$E(C)=Var(C)=Np_i$

But here,

C sample mean = 1.01

C sample var = 1.50

Mixture Model

Suppose the p_i are actually from a distribution.

p_i ~ Gamma(a, b)

C ~ Integral(P(C|p_i) * P(p_i)) dp_i

= Negative Binomial(c, d)

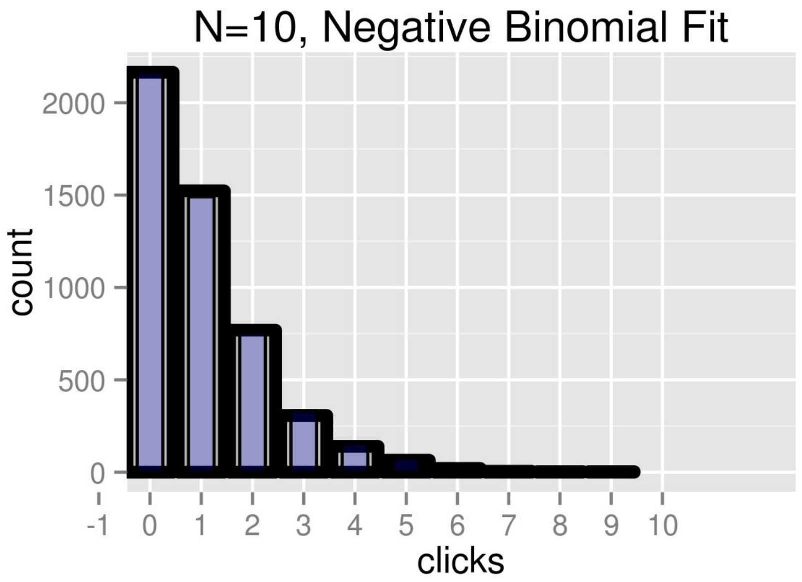
Gamma Mixture of Poisson

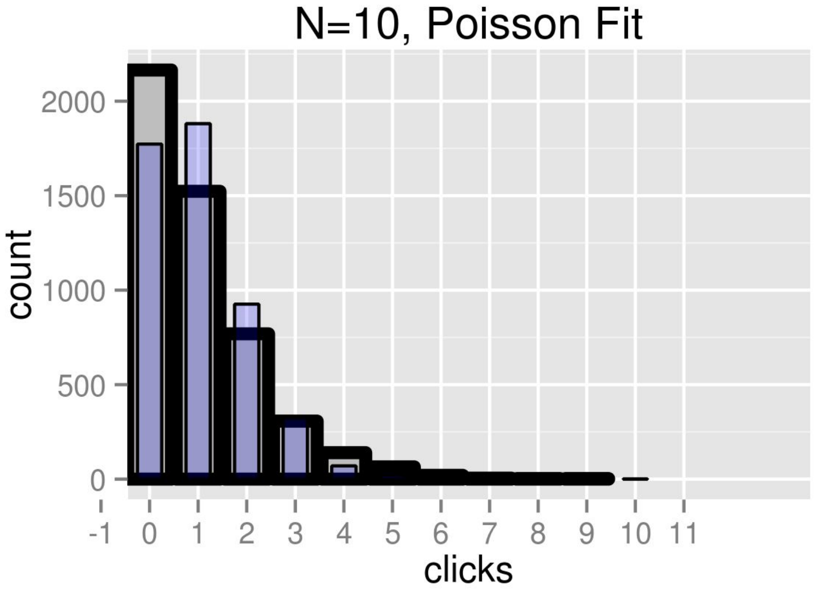
$$
\begin{aligned}
f(k;r,p) &= \int_0^\infty f_{\mathrm{Poisson}(\lambda)}(k)\cdot f_{\mathrm{Gamma}\left(r,\,\frac{p}{1-p}\right)}(\lambda)\,\mathrm{d}\lambda \\
&= \int_0^\infty \frac{\lambda^k}{k!}e^{-\lambda}\cdot\lambda^{r-1}\frac{e^{-\lambda(1-p)/p}}{\left(\frac{p}{1-p}\right)^r\Gamma(r)}\,\mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)}\int_0^\infty \lambda^{r+k-1}e^{-\lambda/p}\,\mathrm{d}\lambda \\
&= \frac{(1-p)^r p^{-r}}{k!\,\Gamma(r)}\,p^{r+k}\,\Gamma(r+k) \\
&= \frac{\Gamma(r+k)}{k!\,\Gamma(r)}\,p^k(1-p)^r.
\end{aligned}
$$

A gamma mixture of poissons is a negative binomial.

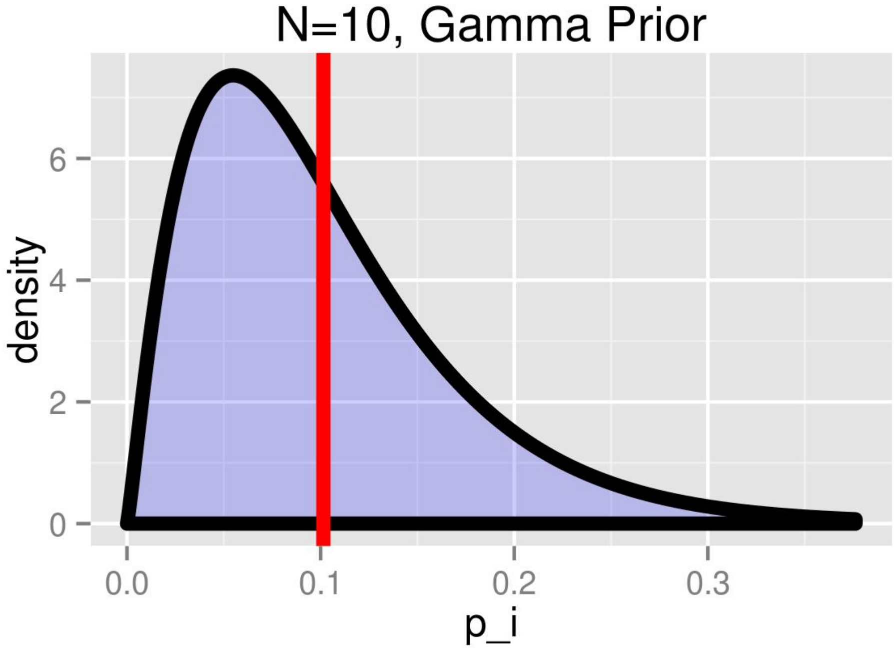
(Formula from Wikipedia)

Negative Binomial Fit





Gamma Prior



Here is what the fit "prior" looks like.

This is an estimate of the distribution of click probabilities based on the overdispersion relative to Poisson.

Bayes' Rule

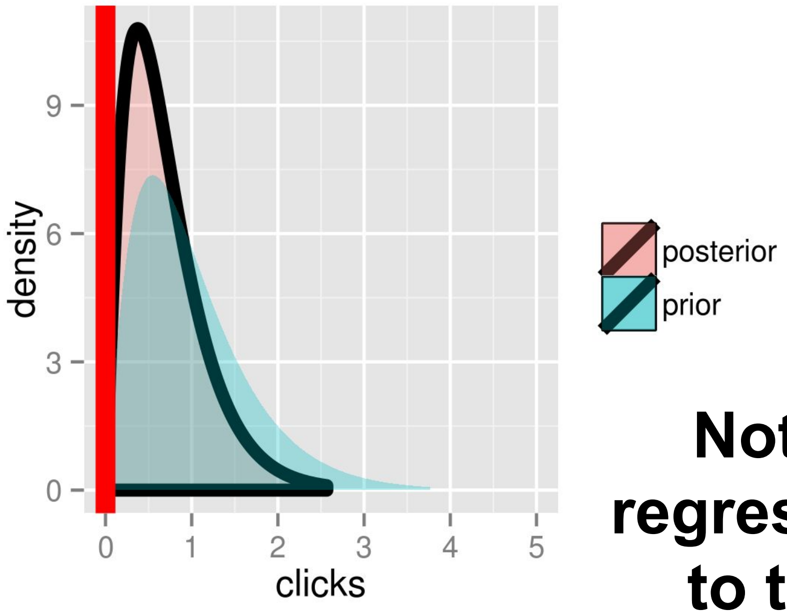
We draw a person from the urn and observe a **C**.

What might their $\mathbf{p_i}$ be?

- We know (have estimated) $\mathbf{P(p_i)}$.
- We know (have assumed) $\mathbf{P(C|p_i)}$.
- What we want is: $\mathbf{P(p_i|C)=P(C|p_i)P(p_i)/P(C)}$

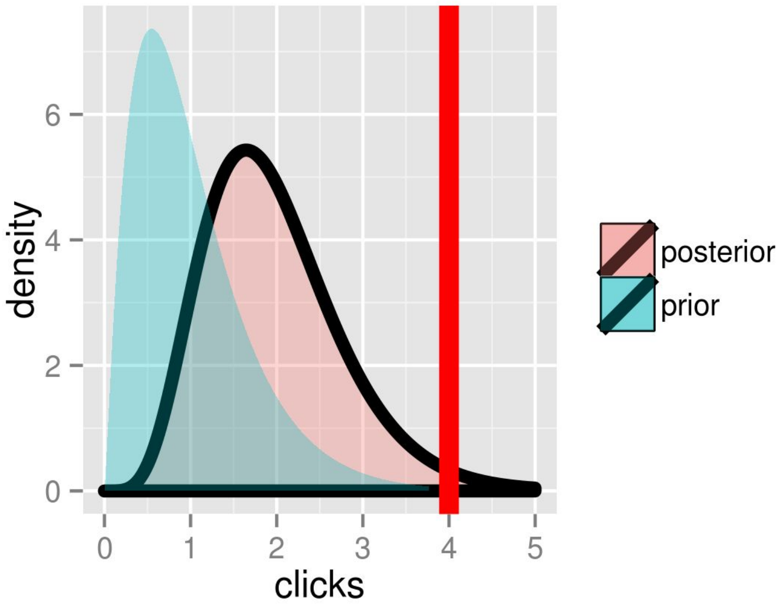
Lucky us! For gamma / poisson, $\mathbf{P(p_i|C)}$ is a gamma distribution (gamma is conjugate for poisson).

Bayes' Rule



Belief about p_i when C=0

Note: regression to the mean



Belief about p_i when C=4

Some Criticism - Meaning

- It is tempting to claim to have measured something about a person. Is this necessarily true? (hint: no)

- How can we test this?

Some Criticism - Meaning

OK	Day 1	Day 2
Person 1	**C=9**	**C=12**
Person 2	C=1	C=2

NOT OK	Day 1	Day 2
Person 1	**C=9**	C=2
Person 2	C=1	**C=12**

A key assumption was that a conditional on *i*, clicks are Poisson.
We knew the noise, so we could detect overdispersion.
If this is violated, EB will mislead you.

Some Criticism - Meaning

- Validate your conditional noise model
  - e.g. here, check that an *individual* is not over-dispersed
- Modeling explicitly can be hard
  - There is generally a nasty integral
- What will you do with it? What is your loss?
  - Sometimes you can measure this empirically.

Some Criticism - The Prior

- Ryan, your choice of prior seems suspiciously convenient.

- What if the fit is not so good?

Some Criticism - The Prior

For certain things, it is enough to fit the marginal:

- e.g. Robbins formula (?), *cf* stat 210 HW3:

$$h(z, \mu) = g(\mu) f(z; \mu) = g(\mu) \exp(z\mu - A(\mu)) f_0(z)$$

$$\mathbb{E}(\mu | Z = z) = \frac{d}{dz}\left(\frac{f}{f_0}\right)(z) \Big/ \left(\frac{f}{f_0}\right)(z) = \frac{d}{dz}\left(\log\left(\frac{f}{f_0}\right)\right)(z).$$

For some things, you are obviously out of luck.

- e.g. does $P(p_i = 0.24572)$ exactly?

Some Criticism - The Prior

Other ways to get priors:

- Point mixtures of conjugate priors (EM)
- Method of moments
- Fit the marginal non-parametrically
  - Deconvolve
  - Use something like Robbins' formula
- Remember, you eventually need a posterior.

Questions Revisited:

What is Bayesian about Empirical Bayes?

When can you use it?

How can you test its assumptions?